Queen's Economics Department Working Paper No. 959

On the Continuity of Correspondences on Sets of Measures with Restricted Marginals

James Bergin
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1997

On the Continuity of Correspondences on Sets of Measures with Restricted Marginals.*

by

J. Bergin

August 1997

Department of Economics
Queen's University
Kingston, Ontario
Canada K7L 3N6

* This paper was written during a visit of the author to the Departament d'Economia i d'Història Econòmica, UAB, Barcelona. The hospitality of the department is gratefully acknowledged.

Abstract

Consider the set of probability measures on a product space with the property that all have the same marginal distributions on the coordinate spaces. This may be viewed as a correspondence, when the marginal distributions are varied. Here, it is shown that this correspondence is continuous. Numerous problems in economics involve optimization over a space of measures where one or more marginal distributions is given. Thus, for this class of problem, Berge's theorem of the maximum is applicable: the set of optimizers is upper-hemicontinuous and the value of the optimal solution varies with the parameters (marginals) continuously.

1 Introduction.

The issue considered here is the (continuity) relationship between joint distributions on product spaces and their marginal distributions. Given marginal distributions on a coordinate space one may associate a set of distributions on the product space with these marginal distributions. This defines a correspondence for marginal to joint distributions. The theorem given below asserts that this correspondence is nonempty, convex-valued and continuous. While the result is of independent interest, there are many game theoretic and economic problems involving optimization where it can usefully be applied.

According to Berge's maximum theorem, if W and Z are topological spaces and $\psi : W \twoheadrightarrow Z$ is a continuous correspondence[1] with nonempty compact values, and $f : W \times Z \to \Re$ is continuous then (1) the function $m(w) = \max_{z \in \psi(w)} f(w, z)$ is continuous, and (2) the correspondence $\beta : X \twoheadrightarrow Y$ defined by $\mu(w) = \{z \in Z \mid f(w, z) = m(w)\}$ is upper-hemicontinuous. There are related results for the cases where f is either an upper or lower semicontinuous function and μ an upper or lower hemicontinuous correspondence. (See Aliprantis and Border (1994) for details.) In the context of this paper, W is a product of sets of probability measures, $w \in W$ denotes a pair of measures (μ, ν) on underlying spaces, X and Y, $z \in Z$ denotes a measure on the product space $X \times Y$ and $\psi(\mu, \nu)$ the set of measures on $X \times Y$ with both marginals agreeing with (μ, ν): if $\tau \in \psi(\mu, \nu)$, then the marginal of τ on X is μ and the marginal on Y is ν. A special case arises when τ is restricted to agree on only one marginal. For the application of Berge's theorem a key question is whether ψ is continuous. Here it is shown that on separable metrizable spaces, the correspondence ψ is continuous. This has a number of useful applications. Some of these are discussed below, following a description of the result. The paper concludes with a proof of the theorem.

2 Framework and Results.

Let X, Y be separable metrizable spaces. A probability measure on $X \times Y$ is denoted τ, while measures on X and Y are denoted μ and ν respectively. Given a set Z, the set of probability measures on Z is denoted $\mathcal{P}(Z)$. If $\tau \in \mathcal{P}(X \times Y)$, let τ_X and τ_Y denote the marginal distributions of τ on X and Y respectively.

Define the correspondence $\psi : \mathcal{P}(X) \times \mathcal{P}(Y) \twoheadrightarrow \mathcal{P}(X \times Y)$ as:

$$\psi(\mu, \nu) = \{\tau \in \mathcal{P}(X \times Y) \mid \tau_X = \mu,\ \tau_Y = \nu\}$$

Thus, $\psi(\mu, \nu)$ is the set of measures on $X \times Y$ whose marginal distributions agree with μ and ν on the coordinate spaces. For example, if $X = Y = [0, 1]$ and $\mu = \nu =$ Lebesgue,

[1] The notation $\psi : W \twoheadrightarrow Z$ denotes a correspondence, while $g : W \to Z$ denotes a function.

then $\tau =$ Lebesgue on $[0,1]^2$ and τ' uniform on $\{(x,y) \mid x = y, \text{ and } x,y \in [0,1]\}$ are both in $\psi(\mu,\nu)$. Since ψ is convex, in this case there are a continuum of distinct measures in $\psi(\mu,\nu)$. The correspondence ψ is nonempty (since $\mu \otimes \nu \in \psi(\mu,\nu)$) and convex (for example, $(\theta\tau + (1-\theta)\tau)_X = \theta\tau_X + (1-\theta)\tau_X$). Below it is shown that ψ is a continuous correspondence. The following remarks identify some implications of the theorem.

Remark 1 Consider the correspondence $\phi(\mu) = \{\tau \in \mathcal{P}(X \times Y) \mid \tau_x = \mu\}$. Note that $\psi(\mu,\nu) \subseteq \phi(\mu)$. The previously mentioned result implies that ϕ is a continuous correspondence. Upper hemicontinuity is clear. To prove lower hemicontinuity, suppose that $\tau \in \phi(\mu)$. Let $\tau_Y = \nu$, so that $\tau \in \psi(\mu,\nu)$. Let $\mu_n \to \mu$ so that $(\mu_n,\nu) \to (\mu,\nu)$. Then, from the theorem, there exists $\{\tau^n\}_{n=1}^{\infty}$, $\tau^n \in \psi(\mu_n,\nu) \subseteq \phi(\mu_n)$ and $\tau^n \to \tau$. Continuity of ϕ in the case where X is compact metric and Y complete, separable and metric is proved in Hopenhayn (1990). A proof for the case where both X and Y are compact is given in Bergin (1996), based on the observation that $\tau = q \otimes \mu$ can be approximated by $q^c \otimes \mu$ where q^c is a continuous function from X to the space of probability measures on Y.

Remark 2 One can consider restrictions of the sort where marginals are approximately restricted. For example, let $\rho(\tau,\psi(\mu,\nu)) = \min_{\tau' \in \psi(\mu,\nu)} d(\tau,\tau')$ where d is a metric on $\mathcal{P}(X \times Y)$ consistent with the weak* topology. For $k > 0$, define

$$\psi_k(\mu,\nu) = \{\tau \in \mathcal{P}(X \times Y) \mid \rho(\tau,\psi(\mu,\nu)) \le k\}$$

Thus, $\psi_k(\mu,\nu)$ is the set of distributions on $X \times Y$ within k of a distribution having marginals exactly equal to (μ,ν). To see that $\psi_k(\mu,\nu)$ is continuous, take $\tau \in \psi_k(\mu,\nu)$ and suppose that $(\mu^n,\nu^n) \to (\mu,\nu)$. Thus, $\rho(\tau,\psi(\mu,\nu)) \le k$. Suppose first that the inequality is strict, so that there is some $\bar\tau \in \psi(\mu,\nu)$, such that $d(\tau,\bar\tau) < k$. Since ψ is continuous, $\exists \tau^n \in \psi(\mu^n,\nu^n)$ and $\tau^n \to \bar\tau$. So, $d(\tau,\tau^n) \le d(\tau,\bar\tau) + d(\bar\tau,\tau^n) < k + d(\bar\tau,\tau^n)$. So, $\tau \in \psi_k(\mu^n,\nu^n)$, $n \ge \bar n$, and hence $\tau \in \psi_k(\mu,\nu)$. Now, suppose that $\tau \in \psi_k(\mu,\nu)$, and $\rho(\tau,\psi(\mu,\nu)) = k$. Let $\tau^j \to \tau$ and $\rho(\tau^j,\psi(\mu,\nu)) < k$. For $n \ge n_j$, $\tau^j \in \psi_k(\mu^n,\nu^n)$. Thus, $\tau^j \in \psi_k(\mu^{n_j},\nu^{n_j})$, $(\mu^{n_j},\nu^{n_j}) \to (\mu,\nu)$ and $\tau^j \to \tau$.

The following discussion describes two important examples.

2.1 Mass Transfer.

The Kantorovich mass transfer problem is the following: Given two measures, μ,ν, on a separable metric space U, let $\psi(\mu,\nu)$ be the set of measures on $U \times U$ with the property that the marginals coincide with μ and ν (i.e. if $\tau \in \psi(\mu,\nu)$ then the marginal on the first coordinate space is μ and ν on the second.) Let $c(x,y)$ be a continuous function on $X \times Y$. The Kantorovich functional is:

$$q(\mu,\nu) = \inf_{\tau \in \psi(\mu,\nu)} \int_{U \times U} c(x,y) d\tau$$

Here, μ is the initial distribution of mass, μ the final distribution, and $\psi(\mu,\nu)$ is the set of admissible plans for transferring mass from one distribution to the other. This problem arises in many contexts (infinite dimensional linear programming, probability theory, information theory and so on). When $\tau \in \psi(\mu,\nu)$, then $\tau = p \otimes \mu$, where p is a conditional probability on Y given X — a measurable function from X to the set of probability measures on Y. Thus, given a Borel set B in Y, $\int P(B,x)\mu(dx)$ is the transfer of mass from X to B. The restriction $\tau_y = \nu$ is the restriction on the class of transfer schemes to $\{p : X \to \mathcal{M}(Y) \mid \nu(B) = \int p(B,x)\mu(dx),\ \forall B \text{ Borel.}\}$. There is a substantial literature characterizing q in various contexts. For example, when c is the metric, d, on U and U is compact,

$$q(\mu,\nu) = \sup_f\{\mid \int_U fd(\mu-\nu);\ \mid f(x) - f(y) \mid \leq d(x,y),\ x,y \in U,\ \sup_{x\in U} \mid f(x) \mid < \infty\}$$

See Rachev (1986) for a survey, and Vajda (1989) for applications in statistics.

In the case where U is complete, the set of measures $\psi(\mu,\nu)$, is tight and since $\psi(\mu,\nu)$ is closed, it is compact. In this case the infimum is attained and Berge's theorem applies: $q(\mu,\nu)$ is a continuous function and $\pi(\mu,\nu) = \arg\min_{\tau\in\psi(\mu,\nu)} \int_{U\times U} c(x,y)d\tau$ is an upper-hemicontinuous correspondence. A specific case of this is the Gini measure of income inequality:

$$G(\mu,\nu) = \inf_{\tau\in\psi(\mu,\nu)} \int_{U\times U} \mid x-y \mid d\tau$$

and from Berge's theorem, $G(\mu,\nu)$ varies continuously with (μ,ν).

2.2 Anonymous Games.

Consider the problem of choosing optimally a joint distribution on a product space where the joint distribution is constrained to have a given marginal distribution on one of the spaces. The problem arises naturally in games with a continuum of players where the fixed marginal distribution gives the distribution over players, and the joint distribution gives the distribution over players and actions (a distributional strategy). Distributional strategies appear widely in the literature. Mas-Colell (1984) gives a discussion of equilibrium in distributional strategies for one-shot games; Jovanovic and Rosenthal (1988) provide analogous results for dynamic games. Bergin and Bernhardt (1992, 1995) extend the work of Jovanovic and Rosenthal to games with aggregate uncertainty. Milgrom and Weber (1986) use distributional strategies in the context of games of incomplete information. For such games, an important question is whether the optimal choice(s) vary continuously with parameters such as the marginal distribution — a question that is particularly important in dynamic models (and answered affirmatively by the result given here). Phrased this way, the result may be viewed as identifying particular circumstances under which Berge's maximum theorem may be applied.

In economic applications, models that make use of the continuum of agents formulation are common. Some examples are Lucas and Prescott (1971), Jovanovic (1982), and Hopenhayn (1990, 1992) where equilibrium coincides with a social planner "surplus" optimization problem. In the surplus optimization problem, a characteristics space S and a choice space X are fixed. Take both to be compact metric spaces. A measure μ on agents characteristics is given, and a distributional strategy is a measure on $X \times S$ with marginal μ on S and the set of such measures is $\mathcal{C}(\mu)$. If surplus at strategy τ is denoted $H(\tau)$, then the surplus maximization problem is: $\max_{\tau \in \mathcal{C}(\mu)} H(\tau)$. Provided $\mathcal{C}(\mu)$ is a *continuous correspondence*, the function $h(\mu) \stackrel{\text{def}}{=} \max_{\tau \in \mathcal{C}(\mu)} H(\tau)$ is continuous in μ and the correspondence $\psi(\mu) \stackrel{\text{def}}{=} \text{argmax}_{\mathcal{C}(\mu)} H(\tau)$ is upper-hemicontinuous in μ. For a multiperiod problem, similar considerations apply. To illustrate, consider the following problem in a two period model. In period $i = 1, 2$, a distributional strategy τ_i (on $X \times S$) determines a payoff of $f_i(\tau, x, s)$ to player s making choice x. Suppose that the state which describes the player evolves stochastically according to a transition probability $P(\cdot \mid x, s)$, continuous in (x, s). A player whose state is s and chooses action x has state characteristic for next period drawn from this distribution. Then, the aggregate distribution next period is given by $\mu'(\cdot) = \int P(\cdot \mid x, s) d\tau$. Define this (continuous) mapping by $\rho(\tau)$. Then, the problem of maximizing total welfare is:

$$\max_{\tau \in C(\mu)} \{ \int f_1(\tau, x, s) d\tau + \max_{\tau' \in C(\rho(\tau))} \int f_2(\tau', x, s) d\tau' \}$$

If $\mathcal{C}$ is continuous, then so is $h_2(\tau) \stackrel{\text{def}}{=} \max_{\tau' \in C(\rho(\tau))} \int f_2(\tau', x, s) d\tau'$ and therefore so also is: $h_1(\tau) \stackrel{\text{def}}{=} \int f_1(\tau, x, s) d\tau + \max_{\tau' \in C(\rho(\tau))} \int f_2(\tau', x, s) d\tau'$. In this case, $\max_{\tau \in C(\mu)} h_1(\tau)$ is a continuous function of μ and the set of maximizers is upper-hemicontinuous. Thus, continuity of $\mathcal{C}$ gives a "well-behaved" problem in the multiperiod setting.

3 The Theorem.

This section gives a statement and proof of the theorem.

Theorem 1 *Let X and Y be separable metrizable spaces and let $\psi : \mathcal{P}(X) \times \mathcal{P}(Y) \twoheadrightarrow \mathcal{P}(X \times Y)$ be defined: $\psi(\mu, \nu) = \{\tau \in \mathcal{P}(X \times Y) \mid \tau_X = \mu, \ \tau_Y = \nu\}$. The correspondence ψ is continuous.*

Proof: Upper-hemi continuity is straightforward to prove. Let f be a bounded continuous function on $X \times Y$. Let $\tau^n \in \psi(\mu^n, \nu^n)$ with $\tau^n \to \tau$ and with $(\mu^n, \nu^n) \to (\mu, \nu)$. Since $\tau^n \to \tau$, $\int_{X \times Y} f d\tau^n \to \int_{X \times Y} f d\tau$. Taking $f(x, y) = g(x)$ where g is any continuous bounded function on X,

$$\int_X g d\mu^n = \int_X g d\tau_X^n = \int_{X \times Y} f d\tau^n \to \int_{X \times Y} f d\tau = \int_X g d\tau_X$$

Since $\int_X g d\mu^n \to \int_X g d\mu$ for each continuous bounded g, $\int_X g d\mu = \int_X g d\tau_X$. Similar reasoning gives $\nu = \tau_Y$, so that $\tau \in \psi(\mu, \nu)$. The remainder of the proof concerns lower hemi-continuity.

Given a separable metrizable topological space Z there is a totally bounded metric d_Z consistent with the topology on Z. On (Z, d_Z), let $C(Z)$ be the set of real valued bounded continuous functions on Z, and let $U(Z)$ be the subset of bounded and uniformly continuous functions on Z. $U(Z)$ is separable. A sequence of measures $\{\tau^n\}$ converges to a measure τ if and only if $\int_Z f d\tau^n \to \int_Z f d\tau$ for every $f \in C(Z)$ (or equivalently, for every $f \in U(Z)$). Because Z is separable and the metrization defining $U(Z)$ totally bounded, a countable dense collection in $U(Z)$ and defines the *weak** topology on $\mathcal{P}(Z)$.

In the present context, since X and Y are separable and metrizable, let d_X and d_Y be totally bounded metrics on X and Y respectively, and let $d_{XY} = d_X + d_Y$ be the totally bounded metric on $Z = X \times Y$. Thus, the set of bounded uniformly continuous functions on $(X \times Y, d_{XY})$ is separable. Therefore, $\tau^n \to \tau$ if and only if, for every $f \in U(X \times Y)$, $\int_{X\times Y} f d\tau^n \to \int_{X\times Y} f d\tau$, or equivalently, for f in a countable dense subset of $U(X \times Y)$. In what follows, it is shown that given $\tau \in \mathcal{P}(X \times Y)$ with $\tau_X = \mu$, $\tau_Y = \nu$, and $(\mu^n, \nu^n) \to (\mu, \nu)$, $\exists \tau^n \in \psi(\mu, \nu)$ and for each $f \in U(X \times Y)$, $\int_{X\times Y} f d\tau^n \to \int_{X\times Y} f d\tau$.

The proof is given in 3 steps: (1) Define a grid on $X \times Y$, (2) On the grid, approximate τ by τ^n, with $\tau^n_X = \mu^n$ and $\tau^n_Y = \nu^n$, (3) Put upper bounds on the distance between τ^n and τ, and using the grid in (1), find a subsequence $\tau^{n_s} \in \psi(\mu^{n_s}, \nu^{n_s})$ such that the distance between τ^{n_s} and τ goes to 0. This completes the proof in view of the fact: If $\psi : W \twoheadrightarrow V$ is a correspondence for W to V, where W and V are first countable, then ψ is lower-hemicontinuous at w if (and only if) $w_n \to w$ and $v \in \psi(w)$ imply that there is a subsequence w_{n_k} with some $v_k \in \psi(w_{n_k})$ for each k and $v_k \to z$.

`Step 1:` Partition X and Y into finite collections of sets $\{X_i\}_{i=1}^k$ and $\{Y_j\}_{j=1}^m$ such that (a) at least $\hat{k} = k - 1$ ($\hat{m} = m - 1$) of the X_i's (Y_j's) have small diameter and the remaining set (if any) has small, μ-measure (ν-measure), and (b) the boundaries of the sets have μ (or ν measure 0).

This step makes use of the following facts. Let W be a topological space. For any subset A, write $\overline{A}$ to denote the closure of A, A° to denote the interior of A, A^c to denote the complement of A in Z and $\partial A = \overline{A} \setminus A^\circ$ to denote the boundary of A. The boundary operation, ∂, satisfies the following properties:

(1) $\forall A \subseteq W$, $\partial A = \partial A^c$,

(2) $\forall\, A,\ B\ \subseteq W$, $\partial(A \cap B) \subseteq \partial A \cup \partial B$,

(3) $\forall\, A,\ B\ \subseteq W$, $\partial(A \cup B) \subseteq \partial A \cup \partial B$, and

(4) If $W = X \times Y$, $\forall\, A \subseteq X$, $\forall\, B \subseteq Y$, $\partial(A \times B) \subseteq (\partial A \times Y) \cup (X \times \partial B)$.

These properties have direct implications for measures of boundaries. For example, (4) implies that if $\mu(\partial A) = 0$ and $\nu(\partial B) = 0$, then $\tau(\partial(A \times B)) \leq \tau(\partial A \times Y) + \tau(X \times \partial B) = \mu(\partial A) + \nu(\partial B) = 0$, where $\mu = \tau_X$ and $\nu = \tau_Y$.

Since X and Y are separable, there are countable dense collection of points $\{x_i\}_{i\in I}$, $I = \{1, 2, \ldots\}$, and $\{y_j\}_{j\in J}$, $J = \{1, 2, \ldots\}$ in X and Y respectively. Let $B_X(x, \delta) = \{y \in X \mid d_X(x, y) < \delta\}$, be a open ball around $x \in X$. Then, given $\underline{\delta} > 0$ for any $\{\delta_i\}_{i\in I}$, $\delta_i \geq \underline{\delta}, \forall i$, $X = \cup_{i\in I} B_X(x_i, \delta_i)$.

Let $\mu \in \mathcal{P}(X)$. Since $\partial B_X(x_i, \delta_i) \subseteq \{y \in X \mid d_X(x_i, y) = \delta_i\}$, $\mu(\partial B_X(x_i, \delta_i)) \leq \mu(\{y \in X \mid d_X(x_i, y) = \delta_i\}$. There are at most a countable number of distinct values for δ_i for which $\mu(\{y \in X \mid d_X(x_i, y) = \delta_i\} > 0$, so, in any interval $[\underline{\delta}, \delta]$, $0 < \underline{\delta} < \delta$, there are a continuum of values for δ_i for which $\mu(\{y \in X \mid d_X(x_i, y) = \delta_i\}) = 0$, and hence $\mu(\partial B_X(x_i, \delta_i)) = 0$. For each i, choose $\delta_i \in [\underline{\delta}, \delta]$ so that $\mu(\partial B_X(x_i, \delta_i)) = 0$. Thus, $\{B_X(x_i, \delta_i)\}_{i\in I}$ satisfies $X = \cup_{i\in I} B_X(x_i, \delta_i)$, $\mu(\partial B_X(x_i, \delta_i)) = 0, \forall i \in I$ and $\underline{\delta} \leq \delta_i \leq \delta$. Define a collection of spheres in Y, $B_Y(y_j, \delta_j)$, similarly, with zero-measure boundaries relative to ν: $Y = \cup_{j\in J} B_Y(y_j, \delta_j)$, $\nu(\partial B_Y(j_j, \delta_j)) = 0, \forall j \in j$ and $\underline{\delta} \leq \delta_j \leq \overline{\delta}$.

Now, let $B_X(r) = \cup_{i=1}^r B_X(x_i, \delta_i)$. Since $B_X(r) \uparrow X$, by continuity of the measure μ, $\mu(B_X(r)) \to \mu(X) = 1$. Thus, for any $\epsilon > 0$, there is some k such that $\mu(B_X(k-1)) \geq 1 - \epsilon$. Let $B_X^i = B_X(x_i, \delta_i)$, $i = 1, \ldots, k-1$, and put $B_X^k = X \setminus B_X(k-1)$ so $\cup_{i=1}^k B_X^i = X$. Similarly, for Y, for some m, the set $B_Y(m-1) = \cup_{j=1}^{m-1} B_Y(y_j, \delta_j)$ satisfies $\nu(B_Y(m-1)) \geq 1 - \epsilon$. Put $B_Y^j = B_Y(y_j, \delta_j)$, $j = 1, \ldots, m-1$, $B_Y^m = B_Y^{m-1+1} = Y \setminus B_Y(m-1)$. From (1), $\partial(\cup_{i=1}^{k-1} B(x_i, \delta_i)) \subseteq \cup_{i=1}^{k-1} \partial B(x_i, \delta_i)$ and $\partial(\cup_{j=1}^{m-1} B(y_j, \delta_j)) \subseteq \cup_{j=1}^{m-1} \partial B(y_j, \delta_j)$. Since $\mu(\partial B(x_i, \delta_i)) = 0$, $\forall i$, $\mu(\partial(\cup_{i=1}^{k-1} B(x_i, \delta_i))) \leq \sum_{i=1}^{k-1} \mu(\partial B(x_i, \delta_i)) = 0$. Similarly, $\nu(\partial(\cup_{j=1}^{m-1} B(y_j, \delta_j))) = 0$. Since $\partial B_X(k-1) = \partial B_X(k-1)^c$, $0 = \mu(\partial B_X(k-1)) = \mu(\partial B_X(k-1)^c) = \mu(\partial(X \setminus B_X(k-1)) = \mu(B_X^k)$. Similarly, $\mu(B_Y^m) = 0$. Thus, the collection $\{B_X^i\}_{i=1}^k$ satisfies $\mu(\partial B_X^i) = 0, \forall i = 1, \ldots, k$ and the collection $\{B_Y^j\}_{j=1}^m$ satisfies $\nu(\partial B_Y^j) = 0, j = 1, \ldots, m$.

Define $X_1 = B_X^1$ so that $\mu(\partial X_1) = 0$. Therefore $\mu(\partial X_1^c) = 0$, since $\partial X_1 = \partial X_1^c$. Set $X_2 = B_X^2 \cap X_1^c$, and observe that $\mu(\partial X_2) \leq \mu(\partial B_X^2) + \mu(\partial X_1^c) = 0$. Let $X(2) = X_1 \cup X_2$, and note that $\mu(\partial X(2)) \leq \mu(\partial X_1) + \mu(\partial X_2) = 0$. Set $X_3 = B_X^3 \cap X(2)^c$. Since $\mu(\partial X(2)) = 0 = \mu(\partial X(2)^c)$, $\mu(\partial X_3) = 0$. Proceed inductively: $X(i) = \cup_{s=1}^i X_s$ and $X_{i+1} = B_X^{i+1} \cap X(i)^c$. This defines $\{X_i\}_{i=1}^k$, with $\cup_{i=1}^k X_i = X$, $X_i \cap X_{i'} = \emptyset$, $i \neq i'$ and $\mu(\partial X_i) = 0, i = 1. \ldots, k$. Define $\{Y_j\}_{j=1}^m$ similarly: $\cup_{j=1}^m Y_i = Y$, $Y_j \cap Y_{j'} = \emptyset$, $j \neq j'$ and $\nu(\partial Y_j) = 0, j = 1, \ldots, m$. By construction, $X_i \subseteq B_X^i = B_X(x_i, \delta_i), i = 1, \ldots, k-1$ with $\delta_i \leq \delta$, so that $\forall x, \tilde{x} \in X_i, i \leq k-1$, $d_X(x, \tilde{x}) \leq 2\delta$. Similarly, for $j = 1, \ldots, m-1$, $y, \tilde{y} \in Y_j$ implies that $d_Y(y, y') \leq 2\delta$. Also, letting $X^* = \cup_{i=1}^{k-1} X_i = B_X(k-1)$, so that $\mu(X^*) \geq 1 - \epsilon$.

Similarly, for $Y^* = \cup_{j=1}^{m-1} Y_j$, $\nu(Y^*) \geq 1-\epsilon$. Then, since $X^* \times Y^* = (X^* \times Y) \cap (X \times Y^*)$,

$$\begin{aligned}
\tau(X^* \times Y^*) &= \tau(X^* \times Y) + \tau(X \times Y^*) - \tau((X^* \times Y) \cup (X \times Y^*)) \\
&= \mu(X^*) + \nu(Y^*) - \tau((X^* \times Y) \cup (X \times Y^*)) \\
&\geq (1-\epsilon) + (1-\epsilon) - 1 \\
&\geq 1 - 2\epsilon
\end{aligned}$$

`Step 2:` In the next step, given (μ^n, ν^n), a measure τ^n is constructed such that (a) $\tau_X^n = \mu^n$, $\tau_Y^n = \nu^n$ and (b) for all i and j, $\tau^n(X_i \times Y_j) = \rho_{ij}^n$ where, as $(\mu^n, \nu^n) \to (\mu, \nu)$, $\rho_{ij}^n \to \tau_{ij} = \tau(X_i \times Y_j)$.

Now, given the measure τ, let $\tau_{ij} = \tau(X_i \times Y_j)$ and let Γ be the matrix with $(i,j)^{th}$ entry τ_{ij}. Let $\mu(X_i) = \mu_i = \sum_j \tau_{ij}$ and $\nu_j(Y_j) = \nu_j = \sum_i \tau_{ij}$. If $\mu_i > 0$, let $\alpha_{ij} = \frac{\tau_{ij}}{\mu_i}$ and if $\mu_i = 0$, let $\alpha_{ij} = \frac{1}{m}$. Thus, $\alpha_{ij}\mu_i = \tau_{ij}$, $\forall i, j$.

Define a matrix, Λ^n:

$$\Lambda^n = \begin{pmatrix} \alpha_{11}\mu_1^n & \alpha_{12}\mu_1^n & \cdots & \alpha_{1m}\mu_1^n \\ \alpha_{21}\mu_2^n & \alpha_{22}\mu_2^n & \cdots & \alpha_{2m}\mu_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{k1}\mu_k^n & \alpha_{k2}\mu_k^n & \cdots & \alpha_{km}\mu_k^n \end{pmatrix}$$

Because $\sum_{j=1}^m \alpha_{ij} = 1$, row i sums to μ_i^n. Because $\mu(\partial X_i) = 0$ for each i, $\mu^n \xrightarrow{\text{weak}^*} \mu$ implies that $\mu_i^n = \mu^n(X_i) \to \mu(X_i) = \mu_i$ for each i, so that $\alpha_{ij} \cdot \mu_i^n \to \alpha_{ij} \cdot \mu_i = \tau_{ij}$. Thus, $\Lambda^n \to \Gamma$ Summing vertically, $\hat{\nu}_j^n = \sum_i \alpha_{ij}\mu_i^n$. Since $\sum_i \alpha_{ij}\mu_i = \sum_i \tau_{ij} = \nu_j$, $\mid \hat{\nu}_j^n - \nu_j \mid \leq \sum_i \alpha_{ij} \mid \mu_i^n - \mu_i \mid$, so that for each j, $\mid \hat{\nu}_j^n - \nu_j \mid \to 0$. Because $\nu^n \to \nu = \tau_Y$, and $\nu(\partial Y_j) = 0$ for each j, $\nu_j^n = \nu^n(Y_j) \to \nu(Y_j) = \nu_j$. Thus for each j, $\mid \hat{\nu}_j^n - \nu_j^n \mid \to 0$.

The following calculations modify the matrix Λ^n to a matrix R^n whose column sums are equal to $\{v_j^j\}_{j=1}^m$ (and whose row sums are $\{\mu_i^n\}_{i=1}^k$). Form this a measure is constructed with marginals μ^n and ν^n and a joint distribution on $X \times Y$ that is approximately τ.

Define $J^+ = \{j \mid \hat{v}_j^n - v_j^n > 0\}$ and $J^- = \{j \mid \hat{v}_j^n - v_j^n < 0\}$. For $j \in J^+$, pick r_j^n such that $r_j^n \hat{v}_j^n = v_j^n$. Thus, $(\hat{v}_j^n - v_j^n) = (1 - r_j^n)\hat{v}_j^n$. Let

$$\beta^n = \sum_{j \in J^+} (\hat{v}_j^n - v_j^n) = -\sum_{j \in J^-} (\hat{v}_j^n - v_j^n) = \sum_{j \in J^+} (1 - r_j^n)\hat{v}_j^n \ .$$

For $j \in J^-$ put $\gamma_j^n = \frac{-(\hat{v}_j^n - v_j^n)}{\beta^n}$. Note that $\sum_{j \in J^-} \gamma_j^n = 1$.

From Λ^n define a matrix, $R^n = \{\rho_{ij}^n\}$, satisfying

1. $\forall (i,j)$, $\mid \alpha_{ij}\mu_i^n - \rho_{ij}^n \mid \to 0$, (as $n \to \infty$)
2. $\forall j$, $\sum_i \rho_{ij}^n = \nu_j^n$, $\forall i$, $\sum_j \rho_{ij}^n = \mu_i^n$,

as follows:

$$\rho_{ij}^n = \begin{cases} r_j^n \alpha_{ij}\mu_i^n & \text{if } j \in J^+, \\ \alpha_{ij}\mu_i^n + \gamma_j^n \sum_{t\in J^+}(1-r_t^n)\alpha_{it}\mu_i^n & \text{if } j \in J^- \\ \alpha_{ij}\mu_i^n & \text{if } j \notin J^+ \cup J^- \end{cases}$$

Thus, for $j \in J^+$, the column sum is: $\sum_{i=1}^k \rho_{ij}^n = r_j^n \sum_{i=1}^k \alpha_{ij}\mu_i^n = r_j^n \hat{\nu}_j^n = \nu_j^n$.

For $j \in J^-$, the column sum is:

$$\begin{aligned}
\sum_{i=1}^k \rho_{ij}^n &= \sum_{i=1}^k [\alpha_{ij}\mu_i^n + \gamma_j^n \sum_{t\in J^+}(1-r_t^n)\alpha_{it}\mu_i^n] \\
&= \sum_{i=1}^k \alpha_{ij}\mu_i^n + \gamma_j^n \sum_{t\in J^+}(1-r_t^n)\sum_{i=1}^k \alpha_{it}\mu_i^n \\
&= \hat{\nu}_j^n + \gamma_j^n \sum_{t\in J^+}(1-r_t^n)\hat{\nu}_t^n \\
&= \hat{\nu}_j^n + \gamma_j^n \beta^n \\
&= \nu_j^n
\end{aligned}$$

Finally, for $j \notin J^+ \cup J^-$, the column sum in R^n and Λ^n is the same and equal to $\nu_j^n (= \hat{\nu}_j^n)$.

Next consider rows. Summing row i of R^n,

$$\begin{aligned}
\sum_{j=1}^m \rho_{ij}^n &= \sum_{j\notin J^+\cup J^-} \rho_{ij}^n + \sum_{j\in J^+} \rho_{ij}^n + \sum_{j\in J^-} \rho_{ij}^n \\
&= \sum_{j\notin J^+\cup J^-} \alpha_{ij}\mu_i^n + \sum_{j\in J^+} r_j^n\alpha_{ij}\mu_i^n + \sum_{j\in J^-} (\alpha_{ij}\mu_i^n + \gamma_j^n \sum_{t\in J^+}(1-r_t^n)\alpha_{it}\mu_i^n) \\
&= \sum_{j\notin J^+\cup J^-} \alpha_{ij}\mu_i^n + \sum_{j\in J^+} r_j^n\alpha_{ij}\mu_i^n + \sum_{j\in J^-} \alpha_{ij}\mu_i^n + [\sum_{j\in J^-}\gamma_j^n] \sum_{t\in J^+}(1-r_t^n)\alpha_{it}\mu_i^n) \\
\\
&= \sum_{j\notin J^+\cup J^-} \alpha_{ij}\mu_i^n + \sum_{j\in J^+} r_j^n\alpha_{ij}\mu_i^n + \sum_{j\in J^-} \alpha_{ij}\mu_i^n + \sum_{t\in J^+}(1-r_t^n)\alpha_{it}\mu_i^n) \\
&= \sum_{j\notin J^+\cup J^-} \alpha_{ij}\mu_i^n + \sum_{j\in J^+} \alpha_{ij}\mu_i^n + \sum_{j\in J^-} \alpha_{ij}\mu_i^n \\
&= \sum_{j=1}^m \alpha_{ij}\mu_i^n \\
&= \mu_i^n
\end{aligned}$$

So, the row sum of row i is μ_i^n.

Define a measure τ_{ij}^n on each rectangle $X_i \times Y_j$ according to $\tau_{ij}^n(B\times C) = \frac{\mu_i^n(B)}{\mu_i^n(X_i)} \cdot \frac{\nu_j^n(C)}{\nu_j^n(Y_j)} \cdot \rho_{ij}^n$, if $\mu_i^n(X_i)\nu_i^n(Y_i) > 0$ and $\tau_{ij}^n(B\times C) = 0$ otherwise. Let τ_{ij}^n be the unique extension to a product measure on $X_i \times Y_j$. Define τ^n on $X \times Y$ according to $\tau^n(B\times C) = \sum_{ij} \tau_{ij}^n(B\times C)$.

Note that

$$\begin{aligned}
\tau^n(B\times Y) &= \sum_j \tau^n_{ij}(B\times Y_j)\\
&= \sum_j \frac{\mu^n_i(B)}{\mu^n_i(X_i)}\cdot\frac{\nu^n_j(Y_j)}{\nu^n_j(Y_j)}\cdot\rho^n_{ij}\\
&= \sum_j \frac{\mu^n_i(B)}{\mu^n_i(X_i)}\cdot\rho^n_{ij}\\
&= \frac{\mu^n_i(B)}{\mu^n_i(X_i)}\sum_j \rho^n_{ij}\\
&= \mu^n_i(B)
\end{aligned}$$

Similarly, $\tau^n(X\times C)=\nu^n_j(C)$. Thus, the measure τ^n has marginals μ^n on X and ν^n on Y. So, $\tau^n\in\psi(\mu^n,\nu^n)$. Also, $\tau^n(X_i\times Y_j)=\rho^n_{ij}$. By construction, as $n\to\infty$, $\rho^n_{ij}\to\tau_{ij}$.

Step 3: For any uniformly continuous function $f: X\times Y\to\Re$, and measures $\tau,\tau'\in\mathcal{M}(X\times Y)$, the following calculations place an upper bound on $\mid\int_{X\times Y} f d\tau-\int_{X\times Y} f d\tau'\mid$ in terms of the partition on step 1, where $\tau_X=\mu$ and $\tau_Y=\nu$. Use this to establish convergence of (the approximating measure) τ^n to τ.

Let $f: X\times Y\to\Re$ be uniformly continuous with $\mid f(x,y)-f(x',y')\mid\le\eta$, whenever $d_X(x,x')<\delta'$ and $d_Y(y,y')<\delta'$. Let $b_f=sup_{(x,y)\in X\times Y}\mid f(x,y)\mid$. Pick $x_i\in X_i$, $i=1,\ldots,k-1$, and $y_j\in Y_j$, $j=1,\ldots,m-1$. Thus, $(x,y)\in X_i\times Y_j$, $i\le k-1, j\le m-1$ implies that $d_X(x,x_i)\le 2\delta$ and $d_Y(y,y_i)\le 2\delta$, so if $\delta=\frac{1}{2}\delta'$, $\mid f(x,y)-f(x_i,y_j)\mid\le\eta$. (Note that δ depends on f and η: $\delta(f,\eta)$.) Define $\bar f$ on $X\times Y$ as $\bar f(x,y)=f(x_i,y_j)$ if $(x,y)\in X_i\times Y_j$, and $\bar f(x,y)=f(x,y)$ if $(x,y)\notin(\cup_{i=1}^{k-1}X_i)\times(\cup_{j=1}^{m-1}Y_j)$. The following calculations place bounds on $\mid\int_{X\times Y} f d\tau-\int_{X\times Y} f d\tau'\mid$.

$$\begin{aligned}
&\mid\int_{X\times Y} f d\tau-\int_{X\times Y} f d\tau'\mid\le\\
&\quad\mid\int_{X\times Y} f d\tau-\int_{X\times Y}\bar f d\tau\mid+\mid\int_{X\times Y}\bar f d\tau-\int_{X\times Y}\bar f d\tau'\mid+\mid\int_{X\times Y}\bar f d\tau'-\int_{X\times Y} f d\tau'\mid
\end{aligned}$$

and consider the three terms separately. For ease of notation, write $Z_{ij}=X_i\times Y_j$ and $Z^*=X^*\times Y^*=\cup_{i=1}^{k-1}\cup_{j=1}^{m-1}X_i\times Y_j$.

$$\begin{aligned}
\mid\int_Z f d\tau-\int_Z\bar f d\tau\mid &= \mid\int_{Z^*} f d\tau-\int_{Z^*}\bar f d\tau\mid+\mid\int_{(Z^*)^c} f d\tau-\int_{(Z^*)^c}\bar f d\tau\mid\\
&= \mid\sum_{i=1}^{k-1}\sum_{j=1}^{m-1}\int_{Z_{ij}} f d\tau-\sum_{i=1}^{k-1}\sum_{j=1}^{m-1}\int_{Z_{ij}}\bar f d\tau\mid+\mid\int_{(Z^*)^c} f d\tau-\int_{(Z^*)^c}\bar f d\tau\mid\\
&\le \sum_{i=1}^{k-1}\sum_{j=1}^{m-1}\int_{Z_{ij}}\mid f-\bar f\mid d\tau+\int_{(Z^*)^c}\mid f-\bar f\mid d\tau\\
&\le \eta\tau(Z^*)+2b_f\tau((Z^*)^c)\\
&\le \eta(1-2\epsilon)+2b_f 2\epsilon
\end{aligned}$$

Where the last two inequalities follow from $\tau(Z^*) \geq 1-2\epsilon$, and $\mid f - \bar{f} \mid \leq \eta$ on $Z_{ij} = X_i \times Y_j$, $i = 1, \ldots k-1$, $j = 1, \ldots m-1$, Thus,

$$\mid \int_{X\times Y} f d\tau - \int_{X\times Y} \bar{f} d\tau \mid \leq (1-2\epsilon)\cdot\eta + 4\epsilon \cdot b_f \tag{1}$$

Next, consider $\mid \int_{X\times Y} \bar{f} d\tau' - \int_{X\times Y} \bar{f} d\tau \mid$,

$$\begin{aligned} \mid \int_{X\times Y} \bar{f} d\tau' - \int_{X\times Y} \bar{f} d\tau \mid &\leq \mid \int_{X^*\times Y^*} \bar{f} d\tau' - \int_{X^*\times Y^*} \bar{f} d\tau \mid \\ &\qquad + \mid \int_{(X^*\times Y^*)^c} \bar{f} d\tau' - \int_{(X^*\times Y^*)^c} \bar{f} d\tau \mid \\ &\leq \sum_{i=1}^{k-1}\sum_{j=1}^{m-1} f(x_{ij}) \mid \tau'(X_i \times Y_j) - \tau(X_i \times Y_j) \mid \\ &\qquad + b_f \cdot [\tau'((X^* \times Y^*)^c) + \tau((X^* \times Y^*)^c)] \end{aligned}$$

So, Let $\Delta_{ij}(\tau, \tau') = \tau'(X_i \times Y_j) - \tau(X_i \times Y_j)$ and for any measurable set A, write $(\tau' + \tau)(A)$ for $\tau'(A) + \tau(A)$. Then,

$$\begin{aligned} &\mid \int_{X\times Y} \bar{f} d\tau' - \int_{X\times Y} \bar{f} d\tau \mid \\ &\qquad \leq \sum_{i=1}^{k-1}\sum_{j=1}^{m-1} f(x_{ij}) \mid \Delta_{ij}(\tau,\tau') \mid + b_f \cdot (\tau + \tau')((X^* \times Y^*)^c) \end{aligned} \tag{2}$$

Finally, the calculations for the third term are similar to the first, except that τ' replaces τ. This gives

$$\mid \int_Z f d\tau' - \int_Z \bar{f} d\tau' \mid \leq \sum_{i=1}^{k-1}\sum_{j=1}^{m-1} \int_{Z_{ij}} \mid f - \bar{f} \mid d\tau' + \int_{(Z^*)^c} \mid f - \bar{f} \mid d\tau'$$

Recalling that $\mid f - \bar{f} \mid \leq \eta$ on Z_{ij},

$$\mid \int_Z f d\tau' - \int_Z \bar{f} d\tau' \mid \leq \eta\tau'(Z^*) + 2b_f\tau'((Z^*)^c) \tag{3}$$

Collecting terms, the bound for $\mid \int_{X\times Y} f d\tau - \int_{X\times Y} f d\tau' \mid$ is:

$$\begin{aligned} \mid \int_{X\times Y} f d\tau - \int_{X\times Y} f d\tau' \mid \leq\ & (1-2\epsilon)\cdot\eta + 4\epsilon\cdot b_f \\ &+ \sum_{i=1}^{k-1}\sum_{j=1}^{m-1} f(x_{ij}) \mid \Delta_{ij}(\tau,\tau') \mid + b_f \cdot (\tau+\tau')((X^* \times Y^*)^c) \\ &+ \eta\tau'((X^* \times Y^*)) + 2b_f\tau'((X^* \times Y^*)^c) \end{aligned}$$

Note that $\partial(X_i \times X_j) \subseteq \partial(X_i \times Y) \cup \partial(X \times Y_j)$, $\tau(\partial(X_i \times X_j)) \leq \mu(\partial X_i) + \nu(\partial Y_j) = 0$ and $\partial(X^* \times Y^*)^c = \partial(X^* \times Y^*)$, $\tau(\partial(X^* \times Y^*)^c) = \tau(\partial(X^* \times Y^*)) = 0$. Thus, if τ' is replaced by τ^n, $\tau^n(X_i \times Y_j) \to \tau(X_i \times Y_j)$, $\forall i, j$ (so $\tau^n(X^* \times Y^*) \to \tau(X^* \times Y^*)$), then $\sum_{i=1}^{k-1} \sum_{j=1}^{m-1} f(x_{ij}) \mid \Delta_{ij}(\tau, \tau^n) \mid \to 0$ and $b_f \cdot (\tau^n + \tau)((X^* \times Y^*)^c) \to b_f \cdot 2\tau((X^* \times Y^*)^c) \leq b_f 2\epsilon$ (since $\tau((X^* \times Y^*)^c) < 2\epsilon$). Then,

$$\begin{aligned} \overline{\lim} \mid \int_{X \times Y} f d\tau^n - \int_{X \times Y} f d\tau \mid &\leq [(1-2\epsilon)\eta + 4\epsilon b_f] + [2\epsilon b_f] + [(1-2\epsilon)\eta + 4\epsilon b_f] \\ &\leq 2(1-2\epsilon)\eta + 10\epsilon b_f \\ &\leq 2\eta + 10\epsilon b_f \end{aligned}$$

Since $X \times Y$ is separable and metrizable, a countable collection of functions, $\{f'_s\}_{s=1}^{\infty}$ in $U(X \times Y)$ may be used to determine convergence in $\mathcal{P}(X \times Y)$. By the normalization $f_s = \frac{1}{2 \cdot |b_{f'_s}|}[f'_s + b_{f'_s}]$, $b_{f'_s} = \sup_{(x,y) \in X \times Y} \mid f'_s(x, y) \mid$, each f_s may be assumed to map into $[0, 1]$. Put $\mathcal{D} = \{f_s\}_{s=1}^{\infty}$. A metric on $\mathcal{P}(X \times Y)$ (yielding the weak* topology) is $d(\tau, \tau') = \sum_{s=1}^{\infty} (\frac{1}{2})^s \mid \int_{X \times Y} f_s d\tau - \int_{X \times Y} f_s d\tau' \mid$. Given ϵ, η and $f_s \in \mathcal{D}$, there is a $\delta(\epsilon, \eta, f_s) > 0$, such that (from step 3):

$$\overline{\lim_n} \mid \int_{X \times Y} f_s d\tau^n - \int_{X \times Y} f_s d\tau \mid \leq 2\eta + 10\epsilon$$

($\sup_{x,y} f_s(x, y) \leq 1$.) For each s, let $\eta_s, \epsilon_s > 0$ and $\eta_s \downarrow 0$, $\epsilon_s \downarrow 0$. Fix $\bar{s}$ and put $\delta_{\bar{s}} = \min\{\delta(\eta_s, f_s) \mid s \leq \bar{s}\}$. For this "grid size", let $\tau_{\bar{s}}^n$ be the measure constructed (with marginals agreeing with (μ^n, ν^n)). Then, $\overline{\lim}_n d(\tau_{\bar{s}}^n, \tau) \leq 2\eta_{\bar{s}} + 10\epsilon_{\bar{s}} + (\frac{1}{2})^{\bar{s}}$. So, for each $\bar{s}$, choose $n(\bar{s})$ so that $d(\tau_{\bar{s}}^{n(\bar{s})}, \tau) \leq 2[\eta_{\bar{s}} + 4\epsilon_{\bar{s}}]$. Thus, $d(\tau_{\bar{s}}^{n(\bar{s})}, \tau) \to 0$. Since $\tau_{\bar{s}}^{n(\bar{s})} \in \psi(\mu^{n(\bar{s})}, \nu^{n(\bar{s})})$, ψ is lower-hemicontinuous at (μ, ν).

This completes the proof.

References

[1] Aliprantis, C. D. and K. C. Border, *Infinite Dimensional Analysis*, Springer-Verlag, Berlin, 1994.

[2] Bergin, J. and D. Bernhardt (1992), "Anonymous Sequential Games with Aggregate Uncertainty", *Journal of Mathematical Economics*, Vol. 21, 543-562.

[3] Bergin, J. and D. Bernhardt (1995), "Anonymous Sequential Games: Existence and Characterization of Equilibria", *Economic Theory*, Vol. 5, 461-489.

[4] Hopenhayn, H. (1992a), "Entry, Exit, and Firm Dynamics in Long Run Equilibrium", *Econometrica* 60, 1127-1150.

[5] Hopenhayn, H. (1990), "Industry equilibrium dynamics: A general competitive theory", *mimeo*.

[6] Jovanovic, B. (1982), "Selection and the Evolution of Industry", *Econometrica* 50, pp 649-670.

[7] Jovanovic, B. and R. W. Rosenthal (1988), "Anonymous Sequential Games", *Journal of Mathematical Economics*, Vol. 17, 77-88.

[8] Lucas, R. and E. Prescott, (1971), "Investment Under Uncertainty", *Econometrica*, 39, 659–681.

[9] Mas-Colell, A. (1984), "On a Theorem of Schmeidler", *Journal of Mathematical Economics*, vol 13, 201-206.

[10] Milgrom P. and R. Weber (1986), "Distributional Strategies for Games with Incomplete Information", *Mathematics of Operations Research*, 10, 619-632.

[11] Rachev, S. T. (1986), "The Monge-Kantorovich Mass Transference Problem and its Stochastic Applications", *Theory of Probability and its Applications*, XXIX, No. 4, 647-676.

[12] Vadja, Igor., (1989), *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, Dordrecht.